

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: HOPWOOD, JOHN JOSEPH; SCOTT, HAMISH STEELE;
WEBER, BIRGIT; BLANCH, LIANNE; ANSON, DONALD STEWART

(ii) TITLE OF INVENTION: SYNTHETIC MAMMALIAN
 α -N-ACETYLGUCOSAMINIDASE AND GENETIC SEQUENCES ENCODING SAME

(iii) NUMBER OF SEQUENCES: 6

(iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: NIXON PEABODY LLP
(B) STREET: 990 STEWART AVENUE
(C) CITY: GARDEN CITY
(D) STATE: NEW YORK
(E) COUNTRY: UNITED STATES
(F) ZIP: 11530

(v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.25

(vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER: 09/077,354
(B) FILING DATE: 22-APRIL-1999

(vii) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER: PCT/US96/00747
(B) FILING DATE: 22-NOV-1996

(viii) ATTORNEY/AGENT INFORMATION:

(A) NAME: POKALSKY, ANN R.
(B) REGISTRATION NUMBER: 34,697
(C) REFERENCE/DOCKET NUMBER: 2249/104

(ix) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: 516 742 4343
(B) TELEFAX: 516 742 4366

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2575 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo sapiens
(F) TISSUE TYPE: Peripheral Blood
(G) CELL TYPE: Leukocyte

(ix) FEATURE:

(A) NAME/KEY: CDS
(B) LOCATION: 102..2330

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

```

CCCCGGCTTA GCCTTCGGGT CCACGTGGCC GGAGGCCGCG AGCTGATTGG ACGCGGGCCG      60
CCCCACCCCC TGGCCGTGCG GGGACCCGCA GGACTGAGAC C ATG GAG GCG GTG      113
                                         Met Glu Ala Val
                                         1

GCG GTG GCC GCG GCG GTG GGG GTC CTT CTC CTG GCC GGG GCC GGG GGC      161
Ala Val Ala Ala Ala Val Gly Val Leu Leu Leu Ala Gly Ala Gly Gly
   5              10              15              20

GCG GCA GGC GAC GAG GCC CGG GAG GCG GCG GCC GTG CGG GCG CTC GTG      209
Ala Ala Gly Asp Glu Ala Arg Glu Ala Ala Ala Val Arg Ala Leu Val
              25              30              35

GCC CGG CTG CTG GGG CCA GGC CCC GCG GCC GAC TTC TCC GTG TCG GTG      257
Ala Arg Leu Leu Gly Pro Gly Pro Ala Ala Asp Phe Ser Val Ser Val
              40              45              50

GAG CGC GCT CTG GCT GCC AAG CCG GGC TTG GAC ACC TAC AGC CTG GGC      305
Glu Arg Ala Leu Ala Ala Lys Pro Gly Leu Asp Thr Tyr Ser Leu Gly
   55              60              65

GGC GGC GGC GCG GCG GCG GTG CGG GTG CGC GGC TCC ACG GGC GTG GCG      353
Gly Gly Gly Ala Ala Arg Val Arg Val Arg Gly Ser Thr Gly Val Ala
   70              75              80

GCC GCC GCG GGG CTG CAC CGC TAC CTG CGC GAC TTC TGT GGC TGC CAC      401
Ala Ala Ala Gly Leu His Arg Tyr Leu Arg Asp Phe Cys Gly Cys His
   85              90              95              100

GTG GCC TGG TCC GGC TCT CAG CTG CGC CTG CCG CGG CCA CTG CCA GCC      449
Val Ala Trp Ser Gly Ser Gln Leu Arg Leu Pro Arg Pro Leu Pro Ala
              105              110              115

GTG CCG GGG GAG CTG ACC GAG GCC ACG CCC AAC AGG TAC CGC TAT TAC      497
Val Pro Gly Glu Leu Thr Glu Ala Thr Pro Asn Arg Tyr Arg Tyr Tyr
              120              125              130

CAG AAT GTG TGC ACG CAA AGC TAC TCC TTC GTG TGG TGG GAC TGG GCC      545
Gln Asn Val Cys Thr Gln Ser Tyr Ser Phe Val Trp Trp Asp Trp Ala
              135              140              145

CGC TGG GAG CGA GAG ATA GAC TGG ATG GCG CTG AAT GGC ATC AAC CTG      593
Arg Trp Glu Arg Glu Ile Asp Trp Met Ala Leu Asn Gly Ile Asn Leu
   150              155              160

```

GCA CTG GCC TGG AGC GGC CAG GAG GCC ATC TGG CAG CGG GTG TAC CTG	641
Ala Leu Ala Trp Ser Gly Gln Glu Ala Ile Trp Gln Arg Val Tyr Leu	
165 170 175 180	
GCC TTG GGC CTG ACC CAG GCA GAG ATC AAT GAG TTC TTT ACT GGT CCT	689
Ala Leu Gly Leu Thr Gln Ala Glu Ile Asn Glu Phe Phe Thr Gly Pro	
185 190 195	
GCC TTC CTG GCC TGG GGG CGA ATG GGC AAC CTG CAC ACC TGG GAT GGC	737
Ala Phe Leu Ala Trp Gly Arg Met Gly Asn Leu His Thr Trp Asp Gly	
200 205 210	
CCC CTG CCC CCC TCC TGG CAC ATC AAG CAG CTT TAC CTG CAG CAC CGG	785
Pro Leu Pro Pro Ser Trp His Ile Lys Gln Leu Tyr Leu Gln His Arg	
215 220 225	
GTC CTG GAC CAG ATG CGC TCC TTC GGC ATG ACC CCA GTG CTG CCT GCA	833
Val Leu Asp Gln Met Arg Ser Phe Gly Met Thr Pro Val Leu Pro Ala	
230 235 240	
TTC GCG GGG CAT GTT CCC GAG GCT GTC ACC AGG GTG TTC CCT CAG GTC	881
Phe Ala Gly His Val Pro Glu Ala Val Thr Arg Val Phe Pro Gln Val	
245 250 255 260	
AAT GTC ACG AAG ATG GGC AGT TGG GGC CAC TTT AAC TGT TCC TAC TCC	929
Asn Val Thr Lys Met Gly Ser Trp Gly His Phe Asn Cys Ser Tyr Ser	
265 270 275	
TGC TCC TTC CTT CTG GCT CCG GAA GAC CCC ATA TTC CCC ATC ATC GGG	977
Cys Ser Phe Leu Leu Ala Pro Glu Asp Pro Ile Phe Pro Ile Ile Gly	
280 285 290	
AGC CTC TTC CTG CGA GAG CTG ATC AAA GAG TTT GGC ACA GAC CAC ATC	1025
Ser Leu Phe Leu Arg Glu Leu Ile Lys Glu Phe Gly Thr Asp His Ile	
295 300 305	
TAT GGG GCC GAC ACT TTC AAT GAG ATG CAG CCA CCT TCC TCA GAG CCC	1073
Tyr Gly Ala Asp Thr Phe Asn Glu Met Gln Pro Pro Ser Ser Glu Pro	
310 315 320	
TCC TAC CTT GCC GCA GCC ACC ACT GCC GTC TAT GAG GCC ATG ACT GCA	1121
Ser Tyr Leu Ala Ala Thr Thr Ala Val Tyr Glu Ala Met Thr Ala	
325 330 335 340	
GTG GAT ACT GAG GCT GTG TGG CTG CTC CAA GGC TGG CTC TTC CAG CAC	1169
Val Asp Thr Glu Ala Val Trp Leu Leu Gln Gly Trp Leu Phe Gln His	
345 350 355	
CAG CCG CAG TTC TGG GGG CCC GCC CAG ATC AGG GCT GTG CTG GGA GCT	1217
Gln Pro Gln Phe Trp Gly Pro Ala Gln Ile Arg Ala Val Leu Gly Ala	
360 365 370	
GTG CCC CGT GGC CGC CTC CTG GTT CTG GAC CTG TTT GCT GAG AGC CAG	1265
Val Pro Arg Gly Arg Leu Leu Val Leu Asp Leu Phe Ala Glu Ser Gln	
375 380 385	

CCT GTG TAT ACC CGC ACT GCC TCC TTC CAG GGC CAG CCC TTC ATC TGG	1313
Pro Val Tyr Thr Arg Thr Ala Ser Phe Gln Gly Gln Pro Phe Ile Trp	
390 395 400	
TGC ATG CTG CAC AAC TTT GGG GGA AAC CAT GGT CTT TTT GGA GCC CTA	1361
Cys Met Leu His Asn Phe Gly Gly Asn His Gly Leu Phe Gly Ala Leu	
405 410 415 420	
GAG GCT GTG AAC GGA GGC CCA GAA GCT GCC CGC CTC TTC CCC AAC TCC	1409
Glu Ala Val Asn Gly Gly Pro Glu Ala Ala Arg Leu Phe Pro Asn Ser	
425 430 435	
ACC ATG GTA GGC ACG GGC ATG GCC CCC GAG GGC ATC AGC CAG AAC GAA	1457
Thr Met Val Gly Thr Gly Met Ala Pro Glu Gly Ile Ser Gln Asn Glu	
440 445 450	
GTG GTC TAT TCC CTC ATG GCT GAG CTG GGC TGG CGA AAG GAC CCA GTG	1505
Val Val Tyr Ser Leu Met Ala Glu Leu Gly Trp Arg Lys Asp Pro Val	
455 460 465	
CCA GAT TTG GCA GCC TGG GTG ACC AGC TTT GCC GCC CGG CGG TAT GGG	1553
Pro Asp Leu Ala Ala Trp Val Thr Ser Phe Ala Ala Arg Arg Tyr Gly	
470 475 480	
GTC TCC CAC CCG GAC GCA GGG GCA GCG TGG AGG CTA CTG CTC CGG AGT	1601
Val Ser His Pro Asp Ala Gly Ala Ala Trp Arg Leu Leu Leu Arg Ser	
485 490 495 500	

GTG TAC AAC TGC TCC GGG GAG GCC TGC AGG GGC CAC AAT CGT AGC CCG	1649
Val Tyr Asn Cys Ser Gly Glu Ala Cys Arg Gly His Asn Arg Ser Pro	
505 510 515	
CTG GTC AGG CGG CCG TCC CTA CAG ATG AAT ACC AGC ATC TGG TAC AAC	1697
Leu Val Arg Arg Pro Ser Leu Gln Met Asn Thr Ser Ile Trp Tyr Asn	
520 525 530	
CGA TCT GAT GTG TTT GAG GCC TGG CGG CTG CTG CTC ACA TCT GCT CCC	1745
Arg Ser Asp Val Phe Glu Ala Trp Arg Leu Leu Thr Ser Ala Pro	
535 540 545	
TCC CTG GCC ACC AGC CCC GCC TTC CGC TAC GAC CTG CTG GAC CTC ACT	1793
Ser Leu Ala Thr Ser Pro Ala Phe Arg Tyr Asp Leu Leu Asp Leu Thr	
550 555 560	
CGG CAG GCA GTG CAG GAG CTG GTC AGC TTG TAC TAT GAG GAG GCA AGA	1841
Arg Gln Ala Val Gln Glu Leu Val Ser Leu Tyr Tyr Glu Glu Ala Arg	
565 570 575 580	
AGC GCC TAC CTG AGC AAG GAG CTG GCC TCC CTG TTG AGG GCT GGA GGC	1889
Ser Ala Tyr Leu Ser Lys Glu Leu Ala Ser Leu Leu Arg Ala Gly Gly	
585 590 595	
GTC CTG GCC TAT GAG CTG CTG CCG GCA CTG GAC GAG GTG CTG GCT AGT	1937
Val Leu Ala Tyr Glu Leu Leu Pro Ala Leu Asp Glu Val Leu Ala Ser	
600 605 610	
GAC AGC CGC TTC TTG CTG GGC AGC TGG CTA GAG CAG GCC CGA GCA GCG	1985
Asp Ser Arg Phe Leu Leu Gly Ser Trp Leu Glu Gln Ala Arg Ala Ala	
615 620 625	
GCA GTC AGT GAG GCC GAG GCC GAT TTC TAC GAG CAG AAC AGC CGC TAC	2033
Ala Val Ser Glu Ala Glu Ala Asp Phe Tyr Glu Gln Asn Ser Arg Tyr	
630 635 640	
CAG CTG ACC TTG TGG GGG CCA GAA GGC AAC ATC CTG GAC TAT GCC AAC	2081
Gln Leu Thr Leu Trp Gly Pro Glu Gly Asn Ile Leu Asp Tyr Ala Asn	
645 650 655 660	
AAG CAG CTG GCG GGG TTG GTG GCC AAC TAC TAC ACC CCT CGC TGG CGG	2129
Lys Gln Leu Ala Gly Leu Val Ala Asn Tyr Thr Pro Arg Trp Arg	
665 670 675	
CTT TTC CTG GAG GCG CTG GTT GAC AGT GTG GCC CAG GGC ATC CCT TTC	2177
Leu Phe Leu Glu Ala Leu Val Asp Ser Val Ala Gln Gly Ile Pro Phe	
680 685 690	
CAA CAG CAC CAG TTT GAC AAA AAT GTC TTC CAA CTG GAG CAG GCC TTC	2225
Gln Gln His Gln Phe Asp Lys Asn Val Phe Gln Leu Glu Gln Ala Phe	
695 700 705	
GTT CTC AGC AAG CAG AGG TAC CCC AGC CAG CCG CGA GGA GAC ACT GTG	2273
Val Leu Ser Lys Gln Arg Tyr Pro Ser Gln Pro Arg Gly Asp Thr Val	
710 715 720	

GAC CTG GCC AAG AAG ATC TTC CTC AAA TAT TAC CCC GGC TGG GTG GCC 2321
Asp Leu Ala Lys Lys Ile Phe Leu Lys Tyr Tyr Pro Gly Trp Val Ala
725 730 735 740

GGC TCT TGS TGATAGATTG GCCACCACTG GGCCTTGTTT TCCGCTAATT 2370
Gly Ser Trp

CCAGGGCAGA TTCCAGGGCC CAGAGCTGGA CAGACATCAC AGGATAACCC AGGCCTGGGA 2430

GGAGGCCCCA CGGCCTGCTG GTGGGGTCTG ACCTGGGGGG ATTGGAGGGA AATGACCTGC 2490

CCTCCACCAC CACCCAAAGT GTGGGATTAA AGTACTGTTT TCTTCCACT TAAAAAAAAA 2550

AAAAAAGTCG AGCGCCCGCG AATTC 2575

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 743 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(ix) FEATURE:

- (A) NAME/KEY: Potentially-glycosylated Asn site,
- (B) LOCATION: 261

(ix) FEATURE:

- (A) NAME/KEY: Potentially-glycosylated Asn site,
- (B) LOCATION: 272

(ix) FEATURE:

- (A) NAME/KEY: Potentially-glycosylated Asn site,
- (B) LOCATION: 435

(ix) FEATURE:

- (A) NAME/KEY: Potentially-glycosylated Asn site,
- (B) LOCATION: 503

(ix) FEATURE:

- (A) NAME/KEY: Potentially-glycosylated Asn site,
- (B) LOCATION: 513

(ix) FEATURE:

- (A) NAME/KEY: Potentially-glycosylated Asn site,
- (B) LOCATION: 526

(ix) FEATURE:

- (A) NAME/KEY: Potentially-glycosylated Asn site,
- (B) LOCATION: 532

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Met Glu Ala Val Ala Val Ala Ala Ala Val Gly Val Leu Leu Leu Ala
1 5 10 15

Gly Ala Gly Gly Ala Ala Gly Asp Glu Ala Arg Glu Ala Ala Ala Val
20 25 30

Arg Ala Leu Val Ala Arg Leu Leu Gly Pro Gly Pro Ala Ala Asp Phe
35 40 45

Ser Val Ser Val Glu Arg Ala Leu Ala Ala Lys Pro Gly Leu Asp Thr
50 55 60

Tyr Ser Leu Gly Gly Gly Gly Ala Ala Arg Val Arg Val Arg Gly Ser
65 70 75 80

Thr Gly Val Ala Ala Ala Ala Gly Leu His Arg Tyr Leu Arg Asp Phe
85 90 95

Cys Gly Cys His Val Ala Trp Ser Gly Ser Gln Leu Arg Leu Pro Arg
100 105 110

Pro Leu Pro Ala Val Pro Gly Glu Leu Thr Glu Ala Thr Pro Asn Arg
115 120 125

Tyr Arg Tyr Tyr Gln Asn Val Cys Thr Gln Ser Tyr Ser Phe Val Trp
130 135 140

Trp Asp Trp Ala Arg Trp Glu Arg Glu Ile Asp Trp Met Ala Leu Asn
145 150 155 160

Gly Ile Asn Leu Ala Leu Ala Trp Ser Gly Gln Glu Ala Ile Trp Gln
165 170 175

Arg Val Tyr Leu Ala Leu Gly Leu Thr Gln Ala Glu Ile Asn Glu Phe
180 185 190

Phe Thr Gly Pro Ala Phe Leu Ala Trp Gly Arg Met Gly Asn Leu His
195 200 205

Thr Trp Asp Gly Pro Leu Pro Pro Ser Trp His Ile Lys Gln Leu Tyr
210 215 220

Leu Gln His Arg Val Leu Asp Gln Met Arg Ser Phe Gly Met Thr Pro
225 230 235 240

Val Leu Pro Ala Phe Ala Gly His Val Pro Glu Ala Val Thr Arg Val
245 250 255

Phe Pro Gln Val Asn Val Thr Lys Met Gly Ser Trp Gly His Phe Asn
260 265 270

Cys Ser Tyr Ser Cys Ser Phe Leu Leu Ala Pro Glu Asp Pro Ile Phe
275 280 285

Pro Ile Ile Gly Ser Leu Phe Leu Arg Glu Leu Ile Lys Glu Phe Gly
 290 295 300

Thr Asp His Ile Tyr Gly Ala Asp Thr Phe Asn Glu Met Gln Pro Pro
 305 310 315 320

Ser Ser Glu Pro Ser Tyr Leu Ala Ala Ala Thr Thr Ala Val Tyr Glu
 325 330 335

Ala Met Thr Ala Val Asp Thr Glu Ala Val Trp Leu Leu Gln Gly Trp
 340 345 350

Leu Phe Gln His Gln Pro Gln Phe Trp Gly Pro Ala Gln Ile Arg Ala
 355 360 365

Val Leu Gly Ala Val Pro Arg Gly Arg Leu Leu Val Leu Asp Leu Phe
 370 375 380

Ala Glu Ser Gln Pro Val Tyr Thr Arg Thr Ala Ser Phe Gln Gly Gln
 385 390 395 400

Pro Phe Ile Trp Cys Met Leu His Asn Phe Gly Gly Asn His Gly Leu
 405 410 415

Phe Gly Ala Leu Glu Ala Val Asn Gly Gly Pro Glu Ala Ala Arg Leu
 420 425 430

Phe Pro Asn Ser Thr Met Val Gly Thr Gly Met Ala Pro Glu Gly Ile
 435 440 445

Ser Gln Asn Glu Val Val Tyr Ser Leu Met Ala Glu Leu Gly Trp Arg
 450 455 460

Lys Asp Pro Val Pro Asp Leu Ala Ala Trp Val Thr Ser Phe Ala Ala
 465 470 475 480

Arg Arg Tyr Gly Val Ser His Pro Asp Ala Gly Ala Ala Trp Arg Leu
 485 490 495

Leu Leu Arg Ser Val Tyr Asn Cys Ser Gly Glu Ala Cys Arg Gly His
 500 505 510

Asn Arg Ser Pro Leu Val Arg Arg Pro Ser Leu Gln Met Asn Thr Ser
 515 520 525

Ile Trp Tyr Asn Arg Ser Asp Val Phe Glu Ala Trp Arg Leu Leu Leu
 530 535 540

Thr Ser Ala Pro Ser Leu Ala Thr Ser Pro Ala Phe Arg Tyr Asp Leu
 545 550 555 560

Leu Asp Leu Thr Arg Gln Ala Val Gln Glu Leu Val Ser Leu Tyr Tyr
 565 570 575

Glu Glu Ala Arg Ser Ala Tyr Leu Ser Lys Glu Leu Ala Ser Leu Leu
 580 585 590

```

Arg Ala Gly Gly Val Leu Ala Tyr Glu Leu Leu Pro Ala Leu Asp Glu
    595                                600                                605

Val Leu Ala Ser Asp Ser Arg Phe Leu Leu Gly Ser Trp Leu Glu Gln
    610                                615                                620

Ala Arg Ala Ala Ala Val Ser Glu Ala Glu Ala Asp Phe Tyr Glu Gln
    625                                630                                635                                640

Asn Ser Arg Tyr Gln Leu Thr Leu Trp Gly Pro Glu Gly Asn Ile Leu
    645                                650                                655

Asp Tyr Ala Asn Lys Gln Leu Ala Gly Leu Val Ala Asn Tyr Tyr Thr
    660                                665                                670

Pro Arg Trp Arg Leu Phe Leu Glu Ala Leu Val Asp Ser Val Ala Gln
    675                                680                                685

Gly Ile Pro Phe Gln Gln His Gln Phe Asp Lys Asn Val Phe Gln Leu
    690                                695                                700

Glu Gln Ala Phe Val Leu Ser Lys Gln Arg Tyr Pro Ser Gln Pro Arg
    705                                710                                715                                720

Gly Asp Thr Val Asp Leu Ala Lys Lys Ile Phe Leu Lys Tyr Tyr Pro
    725                                730                                735

Gly Trp Val Ala Gly Ser Trp
    740

```

(2) INFORMATION FOR SEQ ID NO:3:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 10380 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo sapiens
- (viii) POSITION IN GENOME:
 - (A) CHROMOSOME/SEGMENT: Chromosome 17
- (ix) FEATURE:
 - (A) NAME/KEY: exon 1
 - (B) LOCATION: 990..1372
- (ix) FEATURE:
 - (A) NAME/KEY: exon 2
 - (B) LOCATION: 2115..2262

(ix) FEATURE:

(A) NAME/KEY: exon 3
(B) LOCATION: 3056..3202

(ix) FEATURE:

(A) NAME/KEY: exon 4
(B) LOCATION: 3387..3472

(ix) FEATURE:

(A) NAME/KEY: exon 5
(B) LOCATION: 5667..5923

(ix) FEATURE:

(A) NAME/KEY: exon 6
(B) LOCATION: 7745..8955

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

ATAATGAGCA GTGAGGACGA TCAGAGGTCA CCTTCCTGTC TTGGTTTTGG CAGGTTTTGA	60
CCAGTTTCTT TGCTGCATTC TGTTTTATCA GCGGGGTCTT GTGACCTTTT ATCTTGTGCT	120
GACCTCCTGT CTCATCCTGT GACGAAGGCC TAACCTCCTG GGAATTCAGC CCAGCAGGTC	180
TCTGCCTCAT TTTACCCAGC CCCTGTTCAA GATGGAGTCG CTCTGGTTGG AAATCTCTGA	240
CAAATGACA GCTCCTGTGA TGTGCTGCT GCTGCCGCCA ATGGACAGCC TTTAACGTGC	300
CCGCCAGCCC TGCTCCACCG CCGGCCTGGG CTCACATGGC CCCATCCCTC CTCGAACCTC	360
CTAGCCTGTT AGTTACTCAA ATCTGCAAGC TCTCTGCCTT CTCAGGGCCT TCAATAAATG	420
CATTTCTTCT GTCTGGAAGG CTCTTCCTTT CCTCTTCTA GCCAATTCCT ATTCATCCCT	480
GAGTTTCAGA TTAAGAGTCA CTTCCCTTGG AAACCTTACT TCGCTACTTC GCTACTTACT	540
GCACTACTTC GCAGCATCAC AACTATGATG GAAATCCCTA CTTACGTTAA ATATCTGGTT	600
TCTAGGTCAC CTCCTTGACG GGGACGGTAG GGACCGTCTT CTCGTTTCATC AGTAGGGAAG	660
TAGCTATGGC AGTGCCTGAT ACAAATAAAA CTCCTAAATG GTATTATTATTA GATGGTTGGA	720
TGGAAGITAT TTGCGTGTGA AAGCGCGTTT TACCCGAAGG CGCTCTGTGA GGGCCAGCGG	780
GTCCCCCTTC GCCCTGGAGC CGGGGTCACA CGCTCCCCAC CGCGTGCAGT CACGAGACGC	840
CCCCAAGGGA GTATCTCGT ACCCGGAAGC CGCGACTCCT GGCCCTGAGC CCGGGCTTAG	900
CCTTCGGGTC CACGTGGCCG GAGCCGGCAG CTGATTGGAC GCGGGCCGCC CCACCCCTGT	960
GCGGTCGCGG GACCCGCAGG ACTGAGACCA TGGAGGCGGT GCGGTGCGC GCGGCGGTGG	1020
GGGTCTTCTT CCTGGCCGGG GCCGGGGGCG CGGCAGGCGA CGAGGCCCGG GAGGCGGCGG	1080
CCGTGCGGGC GCTCGTGGCC CGGCTGCTGG GGCCAGGCCC CGCGGCCGAC TTCTCCGTGT	1140

CGGTGGAGCG CGCTCTGGCT GCCAAGCCGG GCTTGGACAC CTACAGCCTG GCGCGCGGCG 1200

GCGCGGCGCG CGTGCGGGTG CGCGGCTCCA CGGCGGTGGC GCGCGCCGCG GGGCTGCACC 1260

GCTACCTGCG GCAGTTCTGT GGCTGCCACG TGGCCTGGTC CGGCTCTCAG CTGCGCCTGC 1320

CGCGGCCACT GCCAGCCGTG CCGGGGGAGC TGACCGAGGC CACGCCCAAC AGGTACC GCC 1380

CCGAAGCTTC CCCGCGTCCG CCCGAGGCGC TTACCCCTC CCGGAGCCGC TGCCACCCAA 1440

ATCGGAGGCG TGAGCGGGGA GCGCTGGCCG GAAGGCCACG CTGCGCCGCC TCCAGCAGCT 1500

GTGTGGCCTT GAGCCAGCCA CTCTGCCTTT CAGAGCCTCG GCTGGCCCCC CTGAAAAACG 1560

GAAAGAAGAC GCCTACCGTG CAGTGTATT GTGAGGATTT GCACGATGAT GGGCATAGAA 1620

TTTGTGGTGC ACAATTGGTG ATGAGTGAAT TTTCTTGCCT TCCTCCCCCA CCTTCTCTTT 1680

GAACCTGCGG ACTGAGGAAG GACGCTTCCA TCCCCACCC TACAGGCCTG TGTTCCAGCG 1740

CCTGCCACAC TATGGAGTGA TGTGTTTACA CAGCTGTCTT CCCCTGCCCA TCTGTTAGAC 1800

TGTGGGGGCA GGGATTCCCC GTTCCAGGAA AACACCGTGC AGAGGAGGGG CTCTGGCAGT 1860

GTGGCATGAA AGTGAATAT GCCACCCAAA TACCCGCCAG GCTAGAGGGC CCTGGGAGAG 1920

TGCAGGGGAC GAGTGCCCTA GAAGCCGAGC CCCGGTACCT GGTCTCAGCT CCACCTGGGG 1980

TGGGTCCCAG TGTGCAGCAG AAGGGCCGAG TTTGGAGCCC CTCCCCCTC CTCTAGGTGG 2040

GGGATGGGGG ATTTGTTCCA GGGCCGTGGA CCCTCCAGGG TGGGATGCGC CCCTGTCTAT 2100

GACACTGCCC GCAGGTACCG CTATTACCAG AATGTGTGCA CGCAAAGCTA CTCCTTCGTG 2160

TGGTGGGACT GGGCCCGCTG GGAGCGAGAG ATAGACTGGA TGGCGCTGAA TGGCATCAAC 2220

CTGGCACTGG CCTGGAGCGG CCAGGAGGCC ATCTGGCAGC GGGTGCCTGC CCACTGTCCC 2280

TTCCCCACCC TCCTCTATG CGGGAGCCAC CGTAGGTGTT TTCACCCGCC CCCAGCATG 2340

GGCGCAGTGT CTCTCTTAG AAGTGCTTTC AGCGTGACA GTGGCTTGGG CCTCCTAAAA 2400

ACTGAGGCTT CCGGCCGGGC GCGGTGGCTC ACGCCTGTCA TCCAGCACT TCGGAGGGCC 2460

TAGGCGGGCG GATCAGGAGT TCAGGAGATC GAGACCATCC TGGCCAACAT TGTGAAACCC 2520

CGTCTCTACT AAAATACAAA GAAATAGCAA CCTGGGCAAC AGAGCGAGAC TCTGTCTAAA 2580

AAAAAAAAA AAAAAAAGT AGGCTTCCAG TTTGAGGAGT GGGGCTCCTT CCCCCTCTC 2640

CCCTATGCAG CCAATCACCT GTTCCCTTGG ATCCAATCA TGGGCAGCTC TAGATCTGCC 2700

TCCCTGGAAG CTTCTGTGCT GCAATGGCTG CTCACGGCTC TGCTTAAGCT CTTACACAG 2760

TTGCCCTGCC CTTCCATCTG GCACTCTTGC TCCATGAAGC CTTCTAAGGC CTTCTGTG 2820

GGGGAAAGCC CCTTTGTGCC CCATCTCCTC ACCCATGCGA CAAAGGCAAC ACAGTGAAC 2880

CACCTACTCA CAGGTCTCTT TCCTCTGGGC TGTGGGCTCC TTGATGGCAG CGTTCGGATT 2940

TTGTCTCAGT AGCCCTAGCA CCCAGCACAA AGAAGCAATG AGTGAATGTT TGTTGAATGA 3000

ATGAATGAAT GAATGAAGAT GAATATATTT CTATGTGTGG GCCCTTCTTC CTCAGGTGTA 3060

CCTGGCCTTG GGCCTGACCC AGGCAGAGAT CAATGAGTTC TTTACTGGTC CTGCCTTCCT 3120

GGCCTGGGGG CGAATGGGCA ACCTGCACAC CTGGGATGGC CCCCTGCCCC CCTCCTGGCA 3180

CATCAAGCAG CTTTACCTGC AGGTAAAAGG ATGGAAAAGG GAAGGGGCAG AATCGGTGAT 3240

AGATGGTCAT GGGCCAGGA AGGGTGGTAT TAGGCCGGCC CCAGGGCTCT TAACAGAGGC 3300

GGGGGGCTGC GTGTATCTCG GGAGATGAGG GCCTTCTCAT AGGACAGCAG TGCCCATGCT 3360

CACCACCCTT CCTTCTGTTC CTCCAGCACC GGGTCCTGGA CCAGATGCGC TCCTTCGGCA 3420

TGACCCAGT GCTGCCTGCA TTCGCGGGGC ATGTTCCCGA GGCTGTCAAC AGGTGAGGTT 3480

CCGCTCAGCC CTTCCACTTA GCTCAGAGAG GGAATTTTAT TCCCTTCTAG AACATGACTT 3540

AAAAACTTAA GCTCTGGGCC GGGCGCAGTG GCTCACGCCT GTAATCCCAG CACTTTGGGA 3600

GGCCGAGTTG GCGGATCAG CTGAGGTCAG GAGTTCGAGA CCAGCCTGGC CAACATGGTG 3660

AAACCTGTGC TCTACTAAAA ATATAAAAAT TAGCTGGGCA TGGTGGCAGC CGCCTGTAAT 3720

CCCATCTACT TAGGAGGCTG AGACAGGAGA ATTGCTTAAA CCTGGGAGGC AGACGTTGCA 3780

GTGAGTCAAG ATCACGCCAT TGCACTCCAG CCTGGGTGAC GAGCGAAACT CTGTCTCAAA 3840

CAACAAACA AGCTCTGGAC GTAGGCCTGG GTTTGATTTC TGACTCTGCT ACTAATTAGC 3900

TGTGTGACTT CGGGCAGATG ACATGACTGC TCTGTGCCTC AGTTTCTCTA CTTGTAAAAAT 3960

GGGATCTCTA CCCACTTCGC TGTAGGGTTT GTAATTATCT CTCGATCTAT CTGTGACTTT 4020

GCACAGAGTG CTAGCAATG GCAGCCCTTG GGAGTGGCAG CAGGGGTGCT CCAGTGTCCC 4080

TTGTCCCTCC TGTTCTCTG TGCTTCCAG CCATCCTCTC ACATGTGGTT GGGAAAAGTC 4140

TTCAAGGCTC ACCTGAGACC TCCCCTCCTT CAGGAAGCCT TGCTAGTGCC CCGCATGACC 4200

TCCTTTGCAC CTGCTAATGT CTGGCTCCCA TACTCTCGTA GGACTTAATG CATGCCAGTG 4260

GCCTCCCTGC CCGCCTCTTT GCCCCATCA CCAGGTGGCA GGAACATCAG TCATTCAATC 4320

AATAAACTTG GTCCAGCTGT CTGAGGCTGC CAGAACTGGC TGTGCTGGGT CCTGGGAGGC 4380

GGCAAGAAAG GTGCCCAAGG GCTTACCCCT GATAGGAGAG ATATGTTGGC TGAAGGATAC 4440

AATGTGGGGA CAAGGACAGG AATATATGTG GGTTCGCTC TCCTCTGCCG GGAGAGAGGG 4500

GCAGGAAGGG CTCAGGGCAG AGCCAGCCT TGA AAAATGA GTGTTGCTTG GACGGACGCT 4560

TGGCTAATGC TTGTAATCCT AGCGTTTGG GAGGCTGAGG CGTATGGATC ACCTGCGGTC 4620

AGGAGTTAAA GACCAGCCTG GCCAACATGG CGAAACCCCA TCTCTACTAA AAGTACAAAA 4680

ATTAGCCAGG CGTGGTGGCG GGCTCCTGTA ATCCAGCTA CTCGGTAGGC TGAGGCATGA 4740

GAATCTCTTG AAGCCAGGGG CCAGAGACTG CAGTGAGCCG AGATCACACC ACTTCACTCC 4800

AGCCTGGGTG ACAGAGTGAG ACTCCGCTCT AAAAAAAAAA AAAAAAAG GAAAGAAAT 4860

TAAACACCTC ATGTTCTCAC TCATAGTGGG AGTTGAACAA TGAGAACAAC ATGGACACAG 4920

GAAGGGGAAC ATCACACACC GGGGCCCTTC GCGGTGTGGG GGTCAAGGGG AGGAGTAGCA 4980

TTGGGACAGA TACTTAATGC ATGCGGGGCT GAAACCTAG ATGATGGGTT GATGGGTGCA 5040

GCAAACCACC ATGGCACATG TATACCTATG CAACAAACCT GCATGTCTTG CACAGAACTG 5100

AAC TGAAAGT ATAATTA AAA AAAAAAAG AAGCTGGGTG CGGTGGCCCA CACCTGTAAT 5160

CCCAGCACTT TGGGAGGCCG AGACGGGCGG ATCACAAGGT CAGCAGATCG AGACCATCCT 5220

GGCTAACACA GTGAAACTCA GTCTCTACTA AAAATACAAA AAATTAGCCG GGTGTGGTGG 5280

CGGGCACCTG TAGTCCCAGC TACTAGGGAG GCTGAGGCAG GAGAATGGCA TGAACCTGGG 5340

AGGCAGAGCT TGCAGTGAGC TGAGAATGCG CCACTGCACT CCAGCCTGGG GGACAGAGTG 5400

AGACTCTGCC TCAAAAAAAA AAAAAAAG AAAGAAAAAG GAGCGTGCT TGITTCAGGC 5460

CACAGGAAGG GGAGAGATAG TGAAGTTTT TCAGAGAAGG TGGCCAGGGA AGGAGAAGAA 5520

AGGACTGTAG GCAGAGAGCA TAGCCTGTAC AAAGCCATAG AGGCAAGAGA AACCAGGAGC 5580

TGTAGAGAAG TTGGCAAGGC TGTGTAACAC TATGGTGAAC ACTATGGCGC CTTCCATGAA 5640

ATATCTGAGC TTTTGCTCCC CACTAGGGTG TTCCCTCAGG TCAATGTCAC GAAGATGGGC 5700

AGTTGGGGCC ACTTTAACTG TTCCTACTCC TGCTCCTTCC TTCTGGCTCC GGAAGACCCC 5760

ATATTTCCCA TCATCGGGAG CCTCTCCTG CGAGAGCTGA TCAAGAGTT TGGCACAGAC 5820

CACATCTATG GGGCCGACAC TTCAATGAG ATGCAGCCAC CTTCCCTCAGA GCCCTCCTAC 5880

CTTGCCGCAG CCACCACATG CGTCTATGAG GCCATGACTG CAGGTACAGT GCCTGGGTGG 5940

GGTGGGAGAG CCCCCAGAC CCTCAAAAAG AAGGGAGTAG CAGATGTCAG TAGGGGTAGG 6000

CAGAGGGACT GGAATAATGC CTGCCATAA CACACAGTAC TTTATAGTTT ACCAAGCACG 6060

TGTACACATG CGTTGTCTCA GTGAATCCCA CTGTGTTGA GAGGTGAGCT CTGGAAGCCA 6120

ACAACCTGGG TCACACCTCG CGCTCCTATT TCCTGGCCGT GTGACTTATG ACTCATGACC 6180

TCCTTCCCAG TGTCTCGTTT GCTTTTCCTG TAAACTGGGA CTACCTCATA GGTAGAATAA 6240
CGCCTGGCCC AGAGCAAAGG CCCTAAGAG CTAGCTATGA ACAAGGATT TGTTCATCT 6300
CTGCGTGGTT GCTGAAGTAG GCACTGCAGG CAGGAGGTGA GTGGATGTGC CTAAAGGCAC 6360
TAAGTGCACA TCCTGCTACA AAACGTGTA GCCAGGGCTC CTTCTGCCTA CTTAAAGGAG 6420
GAGTGAGACA GAGGGCGCCC AAGTCAGGAA TGACTTAGTG GAGAGGCGTC TGTGTTGGCC 6480
AGGAAGGGAA CAGATCAGCT CAGCCTTTCT TGAGCAGTAC TGCTCCAAGT GTGACCCAAA 6540
ACCAGCAGCA GCAGCAGCAG CAGCCCGAGC TGTGAGATGG CAAATTCTCA GGCCCTACCC 6600
AAGACCTGAA GGAGAAGCTA CATTTTTTTT TTTTTTGAGA CAGATTTCAC TCTGTTGCTG 6660
AGGCTGGAGC ACAGTGGCAC AATCTCATCT CACTGCAACC TTCGTCTCCT AGGTTCAAGC 6720
GATTCTCCTG CCTCAGCCTC CCGAGTAGCT GGGACTATAG GCACCCGCCA CCACGCCCGG 6780
CAATTTTTGT TTGTTTTGAG ATAGAGTCTC GCTCTGTAC CCAGGCTGGA GTGCAGTGGC 6840
ACGATCTCAG TTCACTGCAA CCTCTGCTTC CTGAGTTCAA GCGATTCTCC TGCCCTCAGCC 6900
TCCTGAGTAG CTGGGATTAC AGGCGCCCCC CAACCACACT CGGCTAATT TGTATTTTTT 6960
AGTAGAGACG GGGTTTTGCT ATGTAGGTCA AGCTGGTTTC AAACCTCTGA CCTCAAATGA 7020
TTCGCCCCACT TCAGCCTCCC AAAGTGTGG GATTACAGGT GTGAGCCACC TTGCCTGGCC 7080
AATTTTTGTA TTTTATAGTA AAACAGGTTT CACCATGGTG GCCAGACTGG TCTCAAACCT 7140
CTGACCTCAG GTGAACCTGCC CACCTCAGCC TCCCAAAGTA CTGGTATTAC AGGCGTGTATC 7200
CACTGCGACT GGCCCTTGATT TTGTTTTTGA GACAGAATCT TACTCTGTCT CCCAGACTGG 7260
AGTGCACTGG CACAATCTCA GCTCACTGCA ACTTCTGCCT CATGGGTTCA AGTGATTCTT 7320
GTGCCTCTAC CTCCCAGTA GCCGGGATTA CAGGCACCTG CCATTACGCT AGGCTAATTT 7380
TTGTATTTTT AGTATAGACA GGGTTTTCCC ACATTGGCCA GGCTGGTCTG GAACTCCTGG 7440
GCTCAAGTGA TCCACCTGCT TCAGCCCCCTC AGAGTACTGG GATTATAGGT GTGGGCCACC 7500
ACGCCCATTC AGAAACCTCC ATGTTTTAAG GAGCCCTCTG GGTAACTCTC ATGTTACACC 7560
AAGTGCTGA ACCCTGTCTC GGAGTTTTCA GAGGGACGCG TATGTGCCAC AGAGCGTCCC 7620
GCTGGTGGGG GTCATGGGAA GCCATGACCT GGGATAGACA GTCGTCTGTA GAGTGGGGTG 7680
AACAATCCCT GGGCCCTCTG TTTCATCACT CCTCTTCTCT GTTCCCCCTA CCTCCTGTCC 7740
ACAGTGGATA CTGAGGCTGT GTGGCTGCTC CAAGGCTGGC TCTCCAGCA CCAGCCCGAG 7800
TTCTGGGGGC CCGCCAGAT CAGGGCTGTG CTGGGAGCTG TGCCCGGTG CCGCCTCCTG 7860

GTTCTGGACC TGTTTGCTGA GAGCCAGCCT GTGTATACCC GCACTGCCTC CTCCAGGGC 7920
 CAGCCCTTCA TCTGTTGCAT GCTGCACAAC TTTGGGGGAA ACCATGGTCT TTTTGGAGCC 7980
 CTAGAGGCTG TGAACGGAGG CCCAGAAGCT GCCCGCCTCT TCCCCAACTC CACCATGGTA 8040
 GGCACGGGCA TGGCCCCCGA GGCATCAGC CAGAACGAAG TGCTCTATTC CCTCATGGCT 8100
 GAGCTGGGCT GGCAGAAAGGA CCCAGTGCCA GATTTGGCAG CCTGGGTGAC CAGCTTTGCC 8160
 GCGCCGGGCT ATGGGGTCTC CCACCCGGAC GCAGGGGCAG CGTGGAGGCT ACTGCTCCGG 8220
 AGTGTGTACA ACTGCTCCGG GGAGGCCTGC AGGGGCCACA ATCGTAGGCC GCTGGTCAGG 8280
 CGGCCGTCCT TACAGATGAA TACCAGCATC TGGTACAACC GATCTGATGT GTTTGAGGCC 8340
 TGGCGGCTGC TGCTCACATC TGCTCCCTCC CTGGCCACCA GCCCGCCTT CCCTACGAC 8400
 CTGCTGGACC TCACTCGGCA GGCAGTGCAG GAGCTGGTCA GCTTGTAATA TGAGGAGGCA 8460
 AGAAGCGCCT ACCTGAGCAA GGAGCTGGCC TCCCTGTTGA GGGCTGGAGG CGTCCTGGCC 8520
 TATGAGCTGC TGCCCGCACT GGACGAGGTG CTGGCTAGTG ACAGCCGCTT CTTGCTGGGC 8580
 AGCTGGCTAG AGCAGGCCCC AGCAGCGGCA GTCAGTGAGG CCGAGGCCGA TTTCTACGAG 8640
 CAGAACAGCC GCTACCAGCT GACCTTGTGG GGGCCAGAAG GCAACATCCT GGCATATGCC 8700
 AACAAAGCAG TGGCGGGGTT GGTGGCCAAC TACTACACCC CTCGCTGGCG GCTTTTCTGT 8760
 GAGGCGCTGG TTGACAGTGT GGGCCAGGSC ATCCCTTTCC AACAGCACCA GTTTGACAAA 8820
 AATGTCTTCC AACTGGAGCA GGCCTTCGTT CTCAGCAAGC AGAGGTATCC CAGCCAGCCG 8880
 CGAGGAGACA CTGTGGACCT GGCCAAGAAG ATCTTCCTCA AATATTACCC CGGCTGGGTG 8940
 GCCGCTCTT GGTGATAGAT TCGCCACCAC TGGGCCTTGT TTTCCGCTAA TTCCAGGGCA 9000
 GATTCCAGGG CCCAGAGCTG GACAGACATC ACAGGATAAC CCAGGCCTGG GAGGAGGCC 9060
 CACGGCCTGC TGGTGGGGTC TGACCTGGGG GGATTGGAGG GAAATGACCT GCCCTCCACC 9120
 ACCACCCAAA GTGTGGGATT AAAGTACTGT TTCTTTTCCA CTAAACTGA TGAGTCCCTT 9180
 GGGTCTGTCA AAATGAGAAG GTCAGTGTG CCACGCTTGG GAGGACTCAG GGCTATAGCA 9240
 TGGCCCTGGG GTGGGACCTG TTCTCCCATC CCTTGCCTCA CGTCCCTGTT TTTGTTTGT 9300
 TGTTTGTTTG TGACGGAGCC TTGGTCTGTT GCCCAGGCTT GAGTACAATG GCACAGTCTC 9360
 GGCTCACTGC AACCTCCGCC TCCTGGGTTT AAGCAATTCT TGTGCCCTCAG CCTCCCCGGT 9420
 AGCTGGGACT ATAGGCATGC ACCACCACAC CAGGCTAATT TTTTTTTTTC CAAGATGGAG 9480
 TCTTGCTCTG TCGCCAGGT TGGAGTTTAG TGGCACCATA TTGGTTTACT GCAACCTCTG 9540

CCTCCCGGGT TCAAGCAATT CTCCTGCCTC AGTCTACCAG GGAGTTAGGA CTACGGGCCT 9600
 GTGCCATCAC GCCTGGCTAA TTTTGTATT TTTTCATAGAG ATAAGGTTTC ACCATGTTGG 9660
 CCAGGCTGGT CTTTAATCC TGAAGTCAAG TGATCCACCT GCCTCGGCCT TCCAAAGTGC 9720
 TGGGATTACA GGAGTGAAGC ACCGTGCCCG GCCATGTCTC TCTTTTAACT ACTAATGTTA 9780
 CCCTGACCTT TGAACGTAGA ATGCCCTTCT GTTGCAGGAA AACCTCTTTT CAAACCATGT 9840
 TTGTCTTTTG CTGGCATGCC ACAGCAACAG TCACCAACAC AGAAGACTTC TGTGACCAAA 9900
 TATTTGGAGG ATTTTCCCA CACACACCAA GCAGCAGACA TCAGCTGGGT GTCTCTCAAT 9960
 TCAGTTCCAA TGTAATCAAC CAGAGACAGC ATCAGATCCC ACAGGGTTAG GGTGCAGATC 10020
 CATGAGACCA CCCCCTCCTT CCCAACGGTT ACAAGTCCTG ATCCCTGGAA TTTCTGACTA 10080
 ACTGGCTTCA AGTTGGAGTT CCCATGACCC CCTTCCCTCT TTTGGAGTCA ACTCATTTGC 10140
 GACAGTGACC CACGAAACAC AGGGAAACCC TTATTATGTT TATTGCTTTA TTACAGAGGA 10200
 AAAAAATTTT TTTCTTTCTT TTTTGAGACA GGGTCTCACT CTGTCAATCA GAATGACTGC 10260
 AGTGGCAGGA TCTGGCTCCG TCACCCAGGC TGGAGTGCAG TGGCATGATC TCGGCTCACT 10320
 ACAGCCTCCA TCCCCCCCCA ACCCCACGCC TCAGCGCCCC ACCCCGCAAG TGGCTGGGAC 10380

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(v) FRAGMENT TYPE: N-terminal

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo sapiens

(ix) FEATURE:

(A) NAME/KEY: Modified-site

(B) LOCATION: 10

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Asp Glu Ala Arg Glu Ala Ala Ala Val Arg Ala Leu Val Ala Arg
1 5 10 15

Leu Leu Gly Pro Gly
20

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 18 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(v) FRAGMENT TYPE: N-terminal

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo sapiens

(ix) FEATURE:

(A) NAME/KEY: Modified-site, glycosylated or
phosphorylated, wherein Xaa may be any
amino acid residue, preferably Arg.

(B) LOCATION: 16

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

Lys Pro Gly Leu Asp Thr Tyr Ser Leu Gly Gly Gly Gly Ala Ala Xaa Val
1 5 10 15

Arg

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 15 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(v) FRAGMENT TYPE: internal

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(ix) FEATURE:
(A) NAME/KEY: Modified-site, glycosylated or
phosphorylated, wherein Xaa may be any
amino acid residue, preferably Ala
(B) LOCATION: 12

(ix) FEATURE:
(A) NAME/KEY: Modified-site, glycosylated or
phosphorylated, wherein Xaa may be any
amino acid residue, preferably Ser
(B) LOCATION: 14

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

Trp	Arg	Leu	Leu	Leu	Thr	Ser	Ala	Pro	Ser	Leu	Xaa	Thr	Xaa	Pro
1				5					10					15